

**ANALISIS KLASIFIKASI OPINI *TWEET* PADA MEDIA SOSIAL
TWITTER MENGGUNAKAN *LATENT DIRICHLET ALLOCATION* (LDA)**



SKRIPSI

**Disusun Sebagai Salah Satu Syarat
Untuk Memperoleh Gelar Sarjana Komputer
pada Departemen Ilmu Komputer / Informatika**

Disusun Oleh :

MINNATI FITRIASIH

24010310141011

**DEPARTEMEN ILMU KOMPUTER / INFORMATIKA
FAKULTAS SAINS DAN MATEMATIKA
UNIVERSITAS DIPONEGORO
2017**

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan di bawah ini :

Nama : Minnati Fitriasih

NIM : 24010310141011

Judul : Analisis Klasifikasi Opini *Tweet* Pada Media Sosial Twitter
Menggunakan *Latent Dirichlet Allocation* (LDA)

Dengan ini saya menyatakan bahwa dalam tugas akhir/skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan di dalam daftar pustaka.

Semarang, Februari 2017



Minnati Fitriasih
24010310141011

HALAMAN PENGESAHAN

Nama : Minnati Fitriasih
NIM : 24010310141011
Judul : Analisis Klasifikasi Opini *Tweet* Pada Media Sosial Twitter Menggunakan
Latent Dirichlet Allocation (LDA)

Telah diujikan pada sidang tugas akhir pada tanggal 27 Januari 2017 dan dinyatakan lulus
pada tanggal 27 Januari 2017.

Semarang, Februari 2017

Mengetahui,
Ketua Departemen Ilmu Komputer/Informatika



Drs. Eko Adi Sarwoko, S.Si, M.Cs
NIP. 19801021 200501 1 003

Panitia Penguji Tugas Akhir
Ketua,

Drs. Eko Adi Sarwoko, M.Kom
NIP. 19651107 199203 1 003

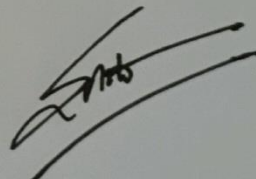
HALAMAN PENGESAHAN

Nama : Minnati Fitriasih
NIM : 24010310141011
Judul : Analisis Klasifikasi Opini *Tweet* Pada Media Sosial Twitter Menggunakan
Latent Dirichlet Allocation (LDA)

Telah diujikan pada sidang tugas akhir pada tanggal 27 Januari 2017 dan dinyatakan lulus pada tanggal 27 Januari 2017.

Semarang, Februari 2017

Dosen Pembimbing,



Dr. Retno Kusumaningrum, S.Si., M.Kom
NIP. 198104202005012001

ABSTRAK

Media sosial menunjukkan perkembangan yang sangat besar pada masa ini. Salah satu platform media sosial yang banyak digunakan di Indonesia adalah Twitter. Penggunaan Twitter yang relatif mudah menjadi pilihan sebagian besar orang dalam menggunakan media sosial ini. Dengan banyaknya pengguna Twitter, semakin banyak informasi yang dihasilkan, tidak jarang pula adanya tindak kejahatan yang dilakukan. Aktivitas kriminal atau tindak kejahatan yang sering dilakukan melalui Twitter pada umumnya berupa *cyber-bullying* dan *cyber-harassment*. Penelitian ini bertujuan untuk mengembangkan aplikasi yang dapat mengklasifikasikan opini *tweet* sebagai *neutral opinion* atau *crime indicated opinion* dengan menggunakan *sentiment analysis*. Opini *tweet* yang akan dikategorikan *crime indicated opinion* yang akan diambil dalam penelitian ini adalah *tweet* atau pesan pada Twitter yang mengandung unsur *cyberbullying* atau *cyberharassment*. *Sentiment analysis* dengan *probabilistic topic model* menggunakan metode *Latent Dirichlet Allocation* (LDA) merupakan metode baru yang dapat diimplementasikan dalam permasalahan analisis klasifikasi opini *tweet*. Hasil penelitian menunjukkan nilai akurasi yang cukup baik sebesar 66% untuk akurasi rata-rata tertinggi dengan nilai hiperparameter β 0,1, α 0,00001 untuk 2 topik pada jumlah iterasi 10000. Berdasarkan hasil akurasi tertinggi, dilakukan perhitungan sensitivitas dan spesifitas untuk setiap *fold* untuk menentukan *fold* terbaik, didapatkan hasil akurasi *fold* tertinggi sebesar 77%.

Kata kunci: *Latent Dirichlet Allocation, Probabilistic Topic Model, Sentiment Analysis, Analisis Klasifikasi Opini Tweet, Twitter*

ABSTRACT

Nowadays, social media showed a great improvement. Twitter is one of social media platforms which is widely used in Indonesia. Simplicity in Twitter usage is the key factor in seizing the wide variety market. Along with the great user statistic, there is malicious activity being performed in twitter. The commonly performed malicious activities are cyber-bullying and cyber harrasment. This research aims to develop applications that can classify tweets opinion as a neutral opinion or crime indicated using sentiment analysis. The tweets that are going to be classified as crime indicated opinion contain cyber-bullying and cyber harrasment content. Sentiment analysis with probabilistic topic model using Latent Dirichlet Allocation(LDA) is a new method which can be implemented in tweet opinion classification analysis. As the result of this study showed a good accuracy score of 66% for average of the highest accuracy with hyperparameter α 0,00001 and β 0,1 for 2 topics at 10000 iteration. Through these result, in order to decide the best fold, a sensitivity and specificity values were calculated over the fold that has highest accuracy. Based on the calculation of sensitivity and specificity for each fold, it was resulted in fold 5 with the highest sensitivity value about 77%.

Keywords: Latent Dirichlet Allocation, Probabilistic Topic Model, Sentiment Analysis, Tweet Opinion Classification Analysis, Twitter

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Allah SWT atas limpahan rahmat dan karunia-Nya penulis dapat menyelesaikan laporan tugas akhir yang berjudul “Analisis Klasifikasi Opini Tweet Pada Media Sosial Twitter Menggunakan *Latent Dirichlet Allocation* (LDA)” dengan baik dan lancar. Laporan tugas akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu (S1) pada Departemen Ilmu Komputer/ Informatika Fakultas Sains dan Matematika Universitas Diponegoro Semarang.

Dalam penyusunan laporan tugas akhir ini, penulis banyak mendapat bimbingan dan dukungan dari berbagai pihak. Oleh karena itu dengan segala kerendahan hati, penulis ingin mengucapkan rasa hormat dan terima kasih kepada :

1. Bapak Ragil Saputra, S.Si, M.Cs, selaku Ketua Departemen Ilmu Komputer / Informatika FSM UNDIP.
2. Ibu Dr. Retno Kusumaningrum, S.Si, M.Kom, selaku dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran dan membantu dalam proses bimbingan hingga terselesaikannya laporan Tugas Akhir ini.
3. Seluruh dosen Departemen Ilmu Komputer/Informatika yang telah memberikan ilmu pengetahuan dan bimbingan.
4. Bapak Firdaus Faishol dan Ibu Istianah, selaku orang tua yang telah sabar dan terus memberikan dukungan moral maupun materi dalam penyelesaian Tugas Akhir ini.
5. Keluarga, Teman – teman mahasiswa Informatika Undip 2010 dan Semua pihak yang telah membantu kelancaran dalam pelaksanaan Tugas Akhir, yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa dalam laporan ini masih banyak terdapat kekurangan baik dari penyampaian materi maupun isi dari materi itu sendiri. Oleh karena itu, kritik dan saran sangat penulis harapkan. Semoga laporan tugas akhir ini dapat bermanfaat bagi penulis dan juga pembaca pada umumnya.

Semarang, Januari 2017

Minnati Fitriasih

DAFTAR ISI

HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	i
HALAMAN PENGESAHAN	ii
HALAMAN PENGESAHAN	iii
ABSTRAK	iv
ABSTRACT	v
KATA PENGANTAR.....	vi
DAFTAR ISI	vii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xii
BAB I PENDAHULUAN	xiii
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan dan Manfaat	3
1.4. Ruang Lingkup.....	4
1.5. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1. Twitter	6
2.2. <i>Sentiment Analysis</i>	7
2.3. <i>Latent Dirichlet Allocation (LDA)</i>	8
2.3.1. Pelatihan dengan <i>Latent Dirichlet Allocation (LDA)</i>	11
2.3.2. Pengujian <i>Latent Dirichlet Allocation (LDA)</i>	13
2.4. <i>Preprocessing</i>	13
2.4.1. Tokenisasi	13
2.4.2. <i>Stemming</i>	14
2.5. <i>K-Fold Cross Validation</i>	22
2.6. <i>Kullback Leiber Divergence (KLD)</i>	23
2.7. Evaluasi.....	24
2.8. Model Pengembangan Aplikasi	25
BAB III METODOLOGI PENELITIAN	27

3.1. Pengumpulan Data	28
3.1.1. Data <i>Neutral Tweet</i>	28
3.1.2. Data <i>Crime Indicated Tweet</i>	29
3.2. <i>Preprocessing</i>	29
3.2.1. Tokenisasi	30
3.2.2. Sastrawi <i>Stemmer</i>	31
3.3. Pembentukan Model Klasifikasi	34
3.3.1. Pelatihan dengan <i>Latent Dirichlet Allocation</i> (LDA).....	34
3.3.2. Pengujian dengan <i>Latent Dirichlet Allocation</i> (LDA).....	47
3.3.3. Evaluasi.....	50
3.4. Pengembangan Aplikasi.....	52
3.4.1. Deskripsi Umum Aplikasi.....	52
3.4.2. Analisis Kebutuhan Aplikasi dan Perancangan Antarmuka Aplikasi.....	53
3.4.3. Implementasi Aplikasi	58
3.4.3.1.Lingkungan Implementasi Aplikasi.....	58
3.4.3.2.Implementasi Antarmuka	58
3.4.3.3. Pengujian Aplikasi.....	60
BAB IV HASIL EKSPERIMEN DAN ANALISIS	61
4.1. Skenario Penelitian.....	61
4.1.1. Skenario 1	62
4.1.2. Skenario 2	62
4.1.3. Skenario 3	62
4.1.4. Skenario 4	62
4.2. Hasil dan Analisis Penelitian	63
4.2.1. Hasil dan Analisis Skenario 1	63
4.2.2. Hasil dan Analisis Skenario 2	65
4.2.3. Hasil dan Analisis Skenario 3	66
4.2.4. Hasil dan Analisis Skenario 4	67
BAB V KESIMPULAN DAN SARAN	70
5.1. Kesimpulan	70
5.2. Saran.....	70
DAFTAR PUSTAKA.....	72

LAMPIRAN-LAMPIRAN	74
-------------------------	----

DAFTAR GAMBAR

Gambar 2.1.	Alur Pembangkitan Teks dengan Topic Model (Liu, 2013)	8
Gambar 2.2.	Topologi LDA (Liu, 2013)	9
Gambar 2.3.	Plate Representation LDA	9
Gambar 2.4.	Ilustrasi LDA sebagai Generatif (Kusumaningrum, 2013).....	10
Gambar 2.5.	Ilustrasi LDA sebagai Proses Inferensi (Kusumaningrum, 2013)	10
Gambar 2.6.	Model LDA (Kusumaningrum et al., 2014).....	11
Gambar 2.7.	Proses <i>10-fold Cross Validation</i>	23
Gambar 2.8.	Model <i>Incremental</i>	25
Gambar 3.1.	Garis Besar Penyelesaian Masalah Klasifikasi Tweet	27
Gambar 3.2.	Flowchart Preprocessing	29
Gambar 3.3.	<i>Flowchart</i> Tokenisasi.....	31
Gambar 3.4.	<i>Flowchart Stemming</i> menggunakan Sastrawi	33
Gambar 3.5.	<i>Flowchart Stemming</i> kata <i>plural</i>	33
Gambar 3.6.	<i>Flowchart stemming</i> kata <i>singular</i>	34
Gambar 3.7.	<i>Flowchart</i> Pelatihan Menggunakan LDA	47
Gambar 3.8.	<i>Flowchart</i> Proses Pengujian	50
Gambar 3.9.	<i>Flowchart</i> Pemilihan Model Klasifikasi Terbaik	52
Gambar 3.10.	<i>Use Case Diagram</i> Analisis Klasifikasi Opini <i>Tweet</i>	54
Gambar 3.11.	<i>Sequence Diagram</i> Mengelola Data Masukan	56
Gambar 3.12.	<i>Sequence Diagram</i> Mengelola Proses Perhitungan.....	56
Gambar 3.13.	<i>Sequence Diagram</i> Menampilkan Hasil Analisis.....	56
Gambar 3.14.	Rancangan Antarmuka Kelola Data Masukan	57
Gambar 3.15.	Rancangan Antarmuka Proses Perhitungan	57
Gambar 3.16.	Rancangan Antarmuka Analisis Klasifikasi Opini <i>Tweet</i>	58
Gambar 3.17.	Implementasi Antarmuka Kelola Data Masukan	59
Gambar 3.18.	Implementasi Antarmuka Perhitungan	59
Gambar 3.19.	Implementasi Antarmuka Analisis Klasifikasi Opini <i>Tweet</i>	59
Gambar 4.1.	Skenario Penelitian	61

Gambar 4.2.	Grafik Perbandingan Rata-rata Pengaruh Jumlah Topik Terhadap Nilai Akurasi.....	64
Gambar 4.3.	Perbandingan Rata-rata Akurasi Terhadap Nilai <i>Alfa</i>	66
Gambar 4.4.	Perbandingan Rata-rata Akurasi Terhadap Nilai <i>Beta</i>	67
Gambar 4.5.	<i>Flowchart</i> Proses Klasifikasi	69

DAFTAR TABEL

Tabel 2.1. Contoh Daftar Fitur pada Twitter	6
Tabel 2.2. Penelitian Terkait <i>Sentiment Analysis</i> Twitter.....	7
Tabel 2.3. Keterangan Notasi dan Definisi Persamaan	12
Tabel 2.4. Contoh Proses Tokenisasi.....	14
Tabel 2.5. Contoh Stemming pada Kata Berbahasa Indonesia.....	15
Tabel 2.6. Daftar Kombinasi Awalan dan Akhiran Yang Tidak Diperbolehkan	16
Tabel 2.7. Aturan Pemenggalan Awalan Algoritma Nazief-Andriani.....	18
Tabel 2.8. Modifikasi dan Tambahan Aturan Pada Algoritma <i>Confix Stripping</i>	19
Tabel 2.9. Modifikasi dan Tambahan Aturan Pada Algoritma ECS	20
Tabel 2.10. Modifikasi Aturan Pemenggalan Awalan dan Penambahan Aturan Pemenggalan Sisipan Pada Modifikasi Algoritma ECS	22
Tabel 2.11. Penambahan Aturan dan Modifikasi Sastrawi <i>Stemmer</i>	22
Tabel 2.12. <i>Confussion Matrix</i> dua kelas	24
Tabel 3.1. Contoh Data <i>Neutral Tweet</i>	28
Tabel 3.2. Contoh Data <i>Tweet Crime Indicated</i>	29
Tabel 3.3. Dataset Sebelum Tokenisasi	30
Tabel 3.4. Dataset Setelah Tokenisasi	30
Tabel 3.5. Kombinasi Awalan dan Akhiran <i>RulePrecedence</i>	32
Tabel 3.6. <i>Bag Of Word</i>	35
Tabel 3.7. Hasil Akhir Proses <i>Gibbs Sampling</i>	42
Tabel 3.8. Hasil Nilai WP	42
Tabel 3.9. Hasil Nilai DP.....	42
Tabel 3.10. Nilai PZD	45
Tabel 3.11. Contoh PZD Uji.....	48
Tabel 3.12. Contoh PZC Latih	48
Tabel 3.13. Tabel <i>Confussion Matrix</i>	51
Tabel 3.14. SRS Aplikasi Analisis Klasifikasi Opini <i>Tweet</i>	53
Tabel 3.15. <i>Use Case Detail</i> Mengelola Data Masukan.....	54
Tabel 3.16. <i>Use Case Detail</i> Mengelola Proses Perhitungan	55

Tabel 3.17. <i>Use Case Detail</i> Menampilkan Hasil Analisis	55
Tabel 4.1. Hasil Akurasi Penelitian	63
Tabel 4.2. Persebaran Kata Pada Jumlah Topik 4	64
Tabel 4.3. Kombinasi dengan Nilai Akurasi Tertinggi Untuk Setiap Nilai <i>Alfa</i>	65
Tabel 4.4. Kombinasi dengan Nilai Akurasi Tertinggi Untuk Setiap Nilai <i>Beta</i>	66
Tabel 4.5. Nilai Akurasi Tertinggi dari Kombinasi Parameter.....	68
Tabel 4.6. Akurasi Fold Kombinasi Terpilih.....	68
Tabel 4.7. Hasil Sensitivitas dan Spesifisitas pada Fold 2, 5, 9.....	68

BAB I

PENDAHULUAN

Bab ini membahas mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, dan sistematika penulisan tugas akhir mengenai Analisis Klasifikasi Opini *Tweet* Pada Media Twitter Menggunakan *Latent Dirichlet Allocation* (LDA).

1.1. Latar Belakang

Media sosial menunjukkan perkembangan yang sangat besar pada masa ini. Hampir sebagian besar orang dari berbagai kalangan menggunakan media sosial dalam kehidupan sehari-hari. Salah satu platform media sosial yang banyak digunakan adalah Twitter. Penggunaan Twitter yang relatif mudah dan tanpa membutuhkan banyak persyaratan untuk mendapatkan satu akun, membuat Twitter menjadi pilihan sebagian besar orang dalam menggunakan media sosial. Twitter adalah layanan pesan sosial gratis untuk mengirim dan menerima pesan secara *real-time*. Ide Twitter pertama kali dicetuskan oleh Jack Dorsey pada tahun 2006. Hingga sekarang ini pengguna Twitter terus bertambah, bahkan pengguna Twitter di Indonesia sendiri mencapai 50 juta orang. Twitter sendiri memiliki beberapa istilah dasar dimana informasi ditampilkan oleh pemilik akun dengan panjang karakter maksimal 140 disebut *tweet* dan penyebaran informasi dari suatu akun oleh akun lain disebut *re-tweet*.

Indonesia menempati urutan kelima negara dengan penggunaan akun Twitter terbanyak (SemioCast, 2012). Sebagai salah satu layanan media sosial yang memiliki jumlah pengguna terbesar di Indonesia, Twitter dapat dikatakan sebagai alat komunikasi modern saat ini. Berbagai aktivitas seperti promosi, aktivitas politik maupun aktivitas kriminal dapat dilakukan melalui Twitter. Persebaran informasi yang sangat cepat dan banyaknya akun pada Twitter juga dapat menyebabkan sulitnya suatu tidak kejahatan tersebut ditemukan. Sebagai ibukota Indonesia, Jakarta menjadi pusat perhatian bagi penduduk Indonesia sendiri. Jakarta mendapatkan urutan pertama dalam kota yang sering melakukan *tweet* (SemioCast, 2012). Dengan

banyaknya penduduk Jakarta dan kebebasan untuk menggunakan media sosial menimbulkan banyaknya aktivitas yang dilakukan melalui Twitter, salah satunya adalah aktivitas kriminal. Sama seperti Jakarta, kota-kota besar di Indonesia juga banyak menggunakan Twitter sebagai salah satu alat komunikasi modern, seperti Bandung yang menempati urutan keenam dunia sebagai salah satu kota di dunia dengan *tweet* terbanyak (SemioCast, 2012).

Aktivitas kriminal atau tindak kejahatan siber yang sering dilakukan melalui Twitter pada umumnya berupa *cyber-bullying* dan *cyber-harassment*. Pada dasarnya *cyber-bullying* dan *cyber-harassment* adalah tindakan mencela atau mempermalukan yang dilakukan tanpa melibatkan tatap muka atau dengan menggunakan perangkat elektronik sebagai penyebaran informasi. Perbedaan utama pada *cyber-bullying* dan *cyber-harassment* terletak pada pelaku dan korban. *Cyber-bullying* ditargetkan untuk minor (anak dan remaja) sebagai pelaku dan korban sedangkan untuk *cyber-harassment* ditargetkan untuk dewasa. Pada penelitian ini, dokumen *tweet* yang diambil diklasifikasikan menjadi kategori *neutral opinion* atau *crime indicated opinion*. *Crime indicated opinion* yang akan diambil dalam penelitian ini adalah *tweet* atau pesan pada Twitter yang mengandung unsur *cyberbullying* atau *cyberharassment*.

Sentiment analysis atau *opinion mining* yang merupakan salah satu bidang studi yang menganalisis opini orang, sentimen, evaluasi, penghargaan, perilaku, dan emosi terhadap suatu produk, jasa, organisasi, individual, topik, atau kejadian tertentu (Liu, 2012). Penelitian mengenai analisis sentimen sudah sering dilakukan. Salah satu penelitian mengenai analisis sentimen yang pernah dilakukan adalah klasifikasi *review* film dengan menggunakan *machine learning* dengan menghasilkan klasifikasi sentimen positif dan negatif (Pang et al, 2002). Penelitian serupa juga dilakukan oleh (Franky & Manurung, 2008), penelitian dilakukan untuk memberikan *review* film dalam bahasa Indonesia. Penelitian ini dilakukan dengan menerjemahkan korpus yang dipakai pada penelitian sebelumnya ke dalam bahasa Indonesia. Pada *sentiment analysis* dataset yang dipakai biasanya sudah terstruktur, koheren dan berbentuk teks dengan minimal panjang teks satu paragraf. Analisis klasifikasi opini *tweet* merupakan *sentiment analysis* pada Twitter. Pada analisis klasifikasi opini *tweet* dilakukan analisis pada *tweet* yang mana berbeda dengan *sentiment analysis* pada umumnya. Pada *tweet*, sentimen disampaikan dalam satu

atau dua kalimat saja, dimana penyampaiannya lebih banyak menggunakan istilah-istilah informal, termasuk pemakaian singkatan dan salah dalam pengetikan (*typo*). (Celykylmaz et al., 2010).

Ada beberapa metode yang digunakan untuk melakukan *sentiment analysis*, salah satu metode yang digunakan adalah *probabilistic topics modeling*. *Probabilistic topics modeling* merupakan salah satu metode yang mampu memberikan hasil yang baik pada bidang *sentiment analysis* (Blei, 2012). *Probabilistic topics modeling* terdiri dari *Probabilistic Latent Semantic Indexing* (PLSA) dan *Latent Dirichlet Allocation* (LDA). PLSA memiliki masalah dalam ketidakkonsistenan pada hasil (Girolami & Kaban, 2003). Oleh karena itu untuk mengatasi masalah pada PLSA dirancang teknik LDA. LDA memberikan kelebihan dalam mengolah data dalam skala besar dan lebih konsisten dibandingkan teknik sebelumnya (Girolami & Kaban, 2003). Pada penelitian ini dilakukan analisis klasifikasi opini *tweet* pada media sosial Twitter menggunakan *Latent Dirichlet Allocation* (LDA).

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, permasalahan yang akan dikaji dalam penelitian tugas akhir ini adalah bagaimana mengembangkan model untuk mengklasifikasi opini *tweet* pada media sosial Twitter sebagai opini yang netral (*neutral opinion*) atau mengarah ke tindak kejahatan (*crime indicated opinion*) dengan metode *Latent Dirichlet Allocation* (LDA).

1.3. Tujuan dan Manfaat

Tujuan yang ingin dicapai dari penelitian tugas akhir ini adalah :

1. Mengetahui pengaruh jumlah topik terhadap nilai akurasi pada pengimplementasian metode LDA untuk analisis klasifikasi opini *tweet*.
2. Mengetahui pengaruh nilai *alfa* terhadap nilai akurasi pada pengimplementasian metode LDA untuk analisis klasifikasi opini *tweet*.

3. Mengetahui pengaruh nilai *beta* terhadap nilai akurasi pada pengimplementasian metode LDA untuk analisis klasifikasi opini *tweet*.
4. Menentukan model klasifikasi terbaik untuk diimplementasikan dalam kasus analisis klasifikasi opini *tweet* untuk klasifikasi opini *tweet* sebagai *neutral opinion* atau *crime indicated opinion*.

Adapun manfaat yang diharapkan dari penelitian ini adalah menghasilkan klasifikasi opini *tweet* pada media sosial Twitter yang nantinya dapat digunakan sebagai referensi atau acuan dalam pengambilan keputusan untuk mengidentifikasi tingkat kejahatan *cyber* (*cyber crime*) di Indonesia.

1.4. Ruang Lingkup

Adapun ruang lingkup dalam pembuatan aplikasi ini adalah sebagai berikut:

1. Metode yang digunakan adalah metode LDA sebagai proses inferensi dengan *collapsed Gibbs Sampling*.
2. Objek data yang diambil berupa *tweet* yang dipilih secara acak dari laman Twitter Indonesia pada bulan Oktober sampai bulan Desember tahun 2015.
3. Objek data diambil berupa *tweet* yang berbahasa Indonesia dengan penulisan menggunakan standar abjad latin.
4. Output yang diharapkan nantinya berupa aplikasi berbasis *web* dengan bahasa pemrograman PHP.

1.5. Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan tugas akhir ini terbagi dalam beberapa pokok bahasan, yaitu:

BAB I PENDAHULUAN

Bab ini memberikan gambaran mengenai latar belakang masalah, rumusan masalah, tujuan dan manfaat, ruang lingkup, serta sistematika penulisan dalam penulisan tugas akhir mengenai Analisis Klasifikasi Opini *Tweet* Pada Media Sosial Twitter Menggunakan *Latent Dirichlet Allocation* (LDA).

BAB II TINJAUAN PUSTAKA

Bab ini membahas kajian pustaka yang digunakan dalam pembuatan tugas akhir. Bab ini menyajikan mengenai Twitter, *Sentimen Analysis*, LDA, model proses yang digunakan dalam pembuatan aplikasi, serta penjelasan mengenai hal-hal terkait dalam penelitian tugas akhir ini.

BAB III METODOLOGI PENELITIAN

Bab ini menyajikan garis besar penyelesaian masalah penyelesaian masalah tugas akhir, proses *preprocessing*, proses pelatihan dan pengujian dengan LDA.

BAB IV HASIL DAN ANALISIS

Bab ini menyajikan hasil penelitian berdasarkan skenario penelitian serta penjelasan mengenai semua hasil dari penelitian dan analisis dari setiap hasil penelitian yang telah dilakukan.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan yang didapatkan dari penelitian dan saran untuk pengembangan penelitian lebih lanjut.